Assessing alexithymia in forensic settings: Psychometric properties of the 20-item Toronto alexithymia scale among incarcerated adult offenders

David A. Preece
Cate L. Parry
Maria [Ricks] M. Allan
Alfred Allan

Follow this and additional works at: https://ro.ecu.edu.au/ecuworkspost2013

Part of the Criminology Commons, and the Psychology Commons

10.1002/cbm.2176
This Journal Article is posted at Research Online. https://ro.ecu.edu.au/ecuworkspost2013/9629
This is the peer reviewed version of the following article: Preece, D. A. Parry, C. L., Allan, M. M., & Allan, A. (2020). Assessing alexithymia in forensic settings: Psychometric properties of the 20-item Toronto alexithymia scale among incarcerated adult offenders. Criminal Behaviour and Mental Health, 31(1), 31-43. https://doi.org/10.1002/cbm.2176, which has been published in final form at https://doi.org/10.1002/cbm.2176. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.
Assessing Alexithymia in Forensic Settings: Psychometric Properties of the 20-item Toronto Alexithymia Scale among Incarcerated Adult Offenders

David A. Preece¹*, Cate L. Parry², Maria M. Allan², Alfred Allan²

¹Curtin University, School of Psychology, Perth, Australia
²Edith Cowan University, School of Arts and Humanities, Perth, Australia

*Corresponding author: David Preece (email: david.preece@curtin.edu.au), Curtin University, School of Psychology, Kent Street, Bentley 6102, Western Australia, Australia.

Acknowledgements: We would like to thank the Department of Justice and Acacia Prison for the use of their sites for data collection. Any material published or made publicly available by researcher cannot be considered as either endorsed by the Department or an expression of the policies or view of the Department. Any errors of omission or commission are the responsibility of the researchers. We also thank the Offender Programs Edith Cowan staff and the prison staff for their assistance during data collection. The data that support the findings of this study are available from the corresponding author upon reasonable request.
Abstract

**Background:** Alexithymia is a trait involving difficulty identifying feelings (DIF), difficulty describing feelings (DDF), and externally orientated thinking (EOT). It is a risk factor for criminal behaviour. It is commonly assessed with the Toronto Alexithymia Scale (TAS-20), but the psychometrics of the TAS-20 have not been tested across the range of offender populations, and it has been suggested it might be unsuitable in incarcerated offenders.

**Aim:** To establish the psychometrics of the TAS-20 among incarcerated offenders.

**Methods:** Factorial validity was examined using confirmatory factor analyses, and the invariance of this factor structure was tested against a published community sample. Reliability coefficients were calculated.

**Results:** 146 incarcerated offenders were recruited. The factor structure of the TAS-20 was invariant across the samples. The intended factor structure composed of DIF, DDF and EOT factors performed well overall (with a reverse-scored method factor added), but six EOT items had low factor loadings. The total scale score and DIF and DDF subscales had acceptable reliability, but EOT did not.

**Conclusions:** Our results suggest that the TAS-20 functions similarly in offender and community samples. Its total scale score, and DIF and DDF subscale scores can be used confidently, but the assessment of EOT may not be adequate with this scale alone. In sum, the TAS-20 can facilitate robust assessments of overall alexithymia in forensic settings.
Assessing Alexithymia in Forensic Settings: Psychometric Properties of the 20-item Toronto Alexithymia Scale in Incarcerated Adult Offenders

The term alexithymia was coined by Sifneos (1973) to describe a cluster of emotion processing deficits often observed in psychiatric patients. It is a multidimensional trait with at least three components: difficulty identifying one’s own feelings (DIF), difficulty describing feelings (DDF), and an externally orientated thinking style (EOT). This last component being characterised by an excessive attentional focus on external stimuli rather than internal emotional states (Nemiah, 1984; Preece et al., 2017; Taylor et al., 1999). Some models of alexithymia also include constricted imaginal processes (IMP) as a fourth component, but this has generally not been supported in factor analytic studies (see Preece et al., 2017, 2020a). Authors of alexithymia questionnaires have, consequently, removed all IMP items from their measures (e.g., Bagby et al., 1994; Sekely et al., 2018a), or exclude them when calculating an alexithymia total score (e.g., Vorst & Bermond, 2001).

Alexithymia is normally distributed in the general population. High levels are an important risk factor for psychiatric disorders characterised by emotional dysregulation (Taylor et al., 1999). Interest in the construct has now extended to forensic settings, with many authors examining alexithymia in offender populations and hypothesising that it might help account for individual differences in criminal behaviour (e.g., Gillespie et al., 2018; Leshem et al., 2019; Payne & Hollin, 2014). Accurate alexithymia assessment is thus important.

Most alexithymia research has so far used the 20-item Toronto Alexithymia Scale (TAS-20; Bagby et al., 1994), which is a self-report measure designed to assess DIF, DDF, and EOT (i.e., in line with a three-component definition of alexithymia). As one of the earliest alexithymia measures developed, it has become influential in the field (Taylor et al.,
2016). To date, however, few studies have examined its psychometric properties in forensic samples, and only two have explored its factor structure (Kroner & Forth, 1995; Parker et al., 2005); that is, established what latent factors or constructs are statistically being tapped by the items. An understanding of its factor structure in this context is important, as it informs knowledge of the theoretical components of alexithymia and dictates what item clusters can be meaningfully used as subscale and total scale scores (Groth-Marnat, 2009). Moreover, TAS-20 scores are often compared between offender and community samples in studies exploring links to criminal behaviour (e.g., Gillespie et al., 2018), but such group comparisons can only be meaningful if its factor structure is demonstrably invariant - that is that items function or are interpreted the same way across the groups (i.e., factorial invariance is supported; Cheung & Rensvold, 2002). So far, no studies have directly tested this. It is thus possible that some observed group differences in TAS-20 scores between forensic and non-forensic samples are explained by measurement effects, rather than true differences in alexithymia. Indeed, one of the TAS-20 authors (Parker et al., 2005) has specifically hypothesised that some TAS-20 items might be less appropriate for incarcerated offenders (e.g., EOT items 16 and 20 that ask about entertainment preferences, where such entertainment preferences may be unavailable to prisoners). If incarcerated offenders also, on average, have lower education levels or reading proficiency than the general community (e.g., Hetland et al., 2007), this might also affect the validity of scores on a self-report measure developed in student/community samples like the TAS-20. The purpose of our study was, therefore, to examine the psychometric properties of the TAS-20 in a sample of incarcerated offenders.

**Methods**

Ethics approval for this project was granted by ethics committees from Edith Cowan University and the Department of Justice (Department) in Western Australia.
The 20-item Toronto Alexithymia Scale

The TAS-20 is a 20-item self-report measure of alexithymia developed by Bagby et al. (1994). Seven items were designed to assess difficulties identifying feelings (DIF subscale), five difficulty describing feelings (DDF subscale) and eight externally orientated thinking (EOT subscale). Items are answered on a 5-point Likert scale, with higher scores generally indicating higher levels of alexithymia; however, five items are reverse-scored. Bagby et al. (1994) designed the scale so that all 20 items could be summed into a total scale score to indicate overall degree of alexithymia, but use/reporting of separate DIF, DDF and EOT subscale scores is also standard practice (e.g., Gillespie et al., 2018; Sekely et al., 2018).

Factorial Validity of the TAS-20. The TAS-20 was developed with adult community, student, and clinical samples. Exploratory factor analysis (EFA; where factors are extracted based on the data, with no a-priori structure imposed) and confirmatory factor analyses (CFA; where theoretically meaningful models are specified and tested) in this early development work provided support for a correlated 3-factor structure (DIF, DDF, EOT) (Bagby et al., 1994). Most subsequent work with similar (non-forensic) adult samples has replicated this theoretically congruent 3-factor model (e.g., Loas et al., 2001; Taylor, Bagby, & Parker, 2003). The fit index values (i.e., indexes from CFAs that indicate how well a specified model fits the data) for this 3-factor model have, however, been mixed, sometimes being below adequate levels, suggesting some potential model misspecifications (see Preece et al., 2018a). Indeed, in all published factor analyses, multiple EOT items have had low factor loadings (<0.40), suggesting that these items may, statistically, not be strong indicators of their intended latent construct; by contrast, all DIF and DDF items usually have strong loadings (i.e., using the commonly used criteria of a factor loading ≥ 0.40 being a meaningful loading; Stevens, 1992).

Authors have usually ascribed this EOT problem to the incorporation of four reverse-
scored items in this subscale. For the TAS-20, as aforementioned, on most items a high score on the 5-point Likert scale indicates high alexithymia (e.g., “It is difficult for me to find the right words for my feelings”), but for the five reverse-scored items their opposite phrasing style means that a high score indicates low alexithymia (e.g., “I am able to describe my feelings easily”). Such item scores must therefore be reversed by the examiner prior to calculating subscale and total scale scores. Reverse-scored items are sometimes inserted in self-report tools as a way to check the validity of self-ratings; if the direction of rating becomes entirely predictable, there is concern that people may almost perseverate. On the other hand, if people have difficulty reading or understanding written questions, this format change may confuse them and decrease response quality. In the general psychometric literature (see van Sonderen et al., 2013), reverse-scored items are often found to have problematic effects, in terms of reducing levels of reliability and creating a ‘method factor’ in the factor structure (i.e., where some variance in these items is accounted for by their unique response format, rather than the substantive latent construct they were designed to measure). This may raise potential content problems with some EOT items (e.g., Gignac et al., 2007; Preece et al., 2018a; Meganck et al., 2008; Kooiman et al., 2002; Tuliao et al., 2019).

Consistent with these views, more recent CFA studies of the TAS-20 have tested factor models with a reverse-scored item method factor added (where all the reverse-scored items were specified to load on this method factor as well as their intended alexithymia factor). This allowed the researcher to partial out the variance attributable to the reverse-scored method effect when looking at whether the items are good indicators of their intended alexithymia factor. Adding the method factor has always improved model fit, usually to adequate levels according to most fit index values (e.g., Preece et al., 2018a; Tuliao et al., 2019; Watters et al., 2016). Some reverse-scored EOT items have loaded more highly on the method factor than the intended EOT factor, indicating that the reverse-scored format may
The two existing CFA studies in forensic samples did not test these more complex (and potentially more accurate) method factor models. One tested only the traditional 3-factor model (Parker et al., 2005) and the other the 3-factor model and an alternate 2-factor model where the DIF and DDF factors were combined (Kroner & Forth, 1995). Kroner and Forth (1995), in their sample of 508 incarcerated offenders, concluded that the 2-factor model (DIF/DDF, EOT) was a better solution. While this solution has also been supported in some non-forensic samples (as DIF and DDF are highly correlated and sometimes extract as a single factor in EFAs; Erni et al., 1997), Parker et al.’s (2005) study with 102 North American Aboriginal incarcerated offenders found that the 3-factor model was adequate. In both cases, though, several EOT items had poor factor loadings (items 5, 15, 16, and 20 in Parker et al., and items 8, 15, 16, and 20 in Kroner & Forth, of which only item 5 is a reverse-scored item). In non-forensic samples, some authors have suggested that a 4-factor model, where EOT is split into separate lack of importance of emotions (IM) and pragmatic thinking (PR) factors might fit better (e.g., Muller et al., 2003); this model has not been tested in forensic samples.

Another limitation of existing forensic CFA studies is that they have not tested any higher-order models – that is those with the first-order factors (e.g., DIF, DDF, EOT) specified to load on a higher-order ‘general alexithymia’ factor. This is problematic, because forensic researchers often use the TAS-20 total scale score (e.g., Velotti et al., 2017), but such a score is only statistically supported if all first-order subscale factors can load together on a higher-order factor (see Brown, 2014). Recent CFAs in non-forensic samples have been encouraging in this respect (e.g., Preece et al., 2018a; Meganck et al., 2008).

**Reliability of the TAS-20.** Studies with non-forensic and forensic samples have found that the internal consistency of the TAS-20 total scale and DIF and DDF subscale scores is
usually acceptable to good, but that the EOT subscale is not (<0.70; for a review, see Kooiman et al., 2002; if the reliability coefficient falls below 0.70, more than 50% of the variance in that score is attributable to error, so the score is judged not suitable for research or clinical/forensic use; e.g., Groth-Marnat, 2009; Preece et al., 2019). In non-forensic samples, splitting the EOT items into IM and PR subscales, or removing all the reverse-scored items, has not improved their reliability (e.g., Preece et al., 2018a; Tuliao et al., 2019), but these alterations have not been tested in forensic samples.

**Concurrent/Criterion Validity of the TAS-20.** Despite the possible problems with the EOT scale, available data suggest that the TAS-20 does measure a variable relevant to criminal behaviour and psychopathology TAS-20 scores have been correlated with other constructs relevant to offending, such as psychopathy (Lander et al., 2012), substance use (Lindsay & Ciarrochi, 2009), impulse control problems (Bibby, 2016), poor social problem solving (Christopher & McMurran, 2009), aggression (Roberton et al., 2014), and empathy deficits (Beadle et al., 2013). Offender samples further usually report significantly higher TAS-20 scores than do community samples (e.g., Strickland et al., 2017). However, as noted earlier, factorial invariance testing is needed to determine whether such score differences between offender and community samples represent true differences in alexithymia or are instead due to the TAS-20 items functioning differently across groups.

**Participants, Procedures, and Materials**

The Department used police records to identify male offenders who were at the time serving sentences in six Western Australia prisons for violent or non-violent crimes. The Department sent out information documents and consent forms to 1687 potential participants and provided a list of those who agreed to participate to the researchers. 146 of the potential participants completed usable TAS-20 questionnaires, giving a response rate of 8.7%. Most participants completed the TAS-20 (English version) in the presence of the first author at
their prison visitor’s centre, but those in regional prisons received and returned it by post.

Following the advice of our cultural consultants that the TAS-20 might not be culturally appropriate for Aboriginal participants, our sample did not include any Aboriginal participants.

**Analytic Strategy**

To maximise our sample size for statistical analysis, like other studies in this area (e.g., Parker et al., 2005), we treated our sample as a general offender sample (rather than separating non-violent and violent offenders). Our sample size ($N=146$) was sufficient for robust factor analyses according to commonly used criteria (5 participants per variable in the analysis or a minimum of 100 participants; Gorsuch, 1983; Kline, 1979).

**Factorial validity.** We conducted confirmatory factor analyses (CFA) using AMOS 25, that is CFAs using maximum likelihood estimation based on a Pearson covariance matrix. Like Preece et al. (2018a), we examined a series of theoretically informed CFA models (see Figure 1).

---Figure 1---

First, we examined several basic first-order correlated models to establish which first-order factor structure best fit our offender sample. These models were: a 1-factor model; a 2-factor correlated model, where the DIF and DDF items were specified to load together on a single DIF/DDF factor and an EOT factor was also included; the traditional 3-factor correlated model, comprised of correlated DIF, DDF, and EOT factors; and a 4-factor correlated model, where EOT was split into separate IM and PR factors.

Model goodness-of-fit was judged based on the pattern of factor loadings and covariances in a model, and four commonly used fit indexes: the Comparative Fit Index
(CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Square Root Mean Square Residual (SRMR). These fit indexes indicate how closely a model fits (or is different from) the patterns in the data. CFI and TLI values ≥0.90 indicate acceptable model fit, and ≥0.95 excellent fit; RMSEA and SRMR values ≤0.08 indicate acceptable fit, and ≤0.06 excellent fit (Byrne, 2010; Marsh et al., 2004). Models were also directly compared using the Akaike Information Criterion (AIC, which penalises model complexity, and lower values indicate a better fitting model; Byrne, 2010) and CFI (a CFI difference >.01 indicates a better fitting model; Cheung & Rensvold, 2002). Item factor loadings of ≥0.40 were judged as meaningful loadings for an item on its specified factor (see Stevens, 1992).

Reverse-scored method factor. Once the best first-order model was determined, we examined this solution with a ‘method’ factor added, whereby all the reverse-scored items (items 4, 5, 10, 18, and 19) were allowed to load together on this method factor, as well as load on their intended alexithymia factor (DDF for item 4 or EOT for items 5, 10, 18, and 19). This allowed us to test how much variance in these item scores was actually measuring an alexithymia component, as opposed to just reflecting the commonality of these items’ response format.

Higher-order factor. We then examined whether the first-order factors from the best solution could load on a higher-order factor, thus supporting the summing of items into a total scale score or an overall marker of alexithymia.

Factorial invariance. After determining the best fitting structure in our offender sample, we examined whether this structure was invariant across our offender sample and a community sample. As the community comparison group, we used a published Australian community sample (N=428) from Preece et al. (2018a); in that sample, the best fitting factor model was the traditional 3-factor correlated model (DIF, DDF, EOT) with a method factor
added. We followed Byrne’s (2010) invariance testing procedure. Firstly, a baseline *configural model* was tested where no equality constraints were imposed across the two samples. Then a *measurement model* was tested with all factor loadings constrained to be equal across the samples. Then a *structural model* was tested with all factor loadings and covariances constrained to be equal across the samples. The factor structure was judged as invariant (i.e., that the items are functioning and being interpreted in the same way across the samples) if the difference in CFI values between the configural model and the measurement or structural models was < 0.01 (see Cheung & Rensvold, 2002).

**Internal consistency.** Using SPSS 25, Cronbach’s α reliability coefficients were calculated for the total scale score and DIF, DDF and EOT subscales. We also calculated these coefficients for the alternative pragmatic thinking and importance of emotions subscales (derived from splitting the EOT subscale; Muller et al., 2003), and for a version of the EOT subscale with the reverse-scored items removed. Reliability coefficients ≥0.70 were judged as acceptable, ≥0.80 as good, and ≥0.90 as excellent (see Groth-Marnat, 2009).

**Results**

**Characteristics of the sample**

Just over half the sample had been convicted of violent (n=79, 54.1%) crimes, and the rest non-violent (n=67, 45.9%) crimes. Nine participants had a small amount of missing data (1 or 2 items) on the TAS-20. These missing data were replaced using the expectation maximisation method (Gold & Bentler, 2000). All TAS-20 items were reasonably normally distributed (max skew=1.44, max kurtosis=-1.58). Age data were provided by 118 of the men, yielding an average age of 37.98 years (SD =12.34, range =19-74). Culturally, 88 (60.3%) reported that they were born in Australia and were non-Aboriginal, 15 (10.3%) were
born outside Australia; 43 (29.5%) did not provide this information.

**Factorial Validity**

Table 1 shows the mean, standard deviations and ranges of the TAS-20 scores for the offender sample. In the first analytic step with the offender sample, we tested the fit of models that did not include a reverse-scored item method factor (i.e., only alexithymia factors). Of these models, the traditional 3-factor model (with correlated difficulty identifying feelings [DIF], difficulty describing feelings [DDF], and externally orientated thinking [EOT] factors) appeared to be the best solution (see Table 2 for fit index values for all models). RMSEA and SRMR fit index values indicated acceptable fit for this model (<.08), however, CFI and TLI fit index values did not reach the acceptable cut-off values (i.e., below .90), thus indicating some potential model problems. All DIF and DDF items loaded well on their intended factor (loadings >0.40), but four EOT items did not (items 5, 16, 18, and 20). Two of these were reverse-scored items. Splitting the EOT factor into importance of emotions and pragmatic thinking factors did not resolve these issues and the 4-factor correlated model did not fit better than the 3-factor correlated model (i.e., CFI difference <0.01). Therefore our preferred solution was the more parsimonious 3-factor correlated model.

Adding a reverse-scored item method factor to the 3-factor model (i.e., DIF, DDF, EOT factors and a method factor) substantially improved levels of fit, to generally adequate levels (see Table 2). The DIF, DDF, and EOT factors in this model were all significantly positively correlated (estimated r-values=0.68-0.83, p-values <0.001). The RMSEA, SRMR and CFI fix index values all indicated acceptable levels of fit, though the TLI fit index values were still slightly below the 0.90 cut-off. With the method factor added, two additional EOT items had low loadings on the EOT factor (thus six EOT items in total with low loadings, with all four reverse-scored items in the EOT subscale being among these items); further, two items loaded more strongly on the method factor than their intended EOT factor (see Table 3)
for all item factor loadings). The EOT factor though, as a whole, did still appear to assess a construct relevant to alexithymia, as a higher-order version of this model (i.e., 3-factor higher-order model+method) indicated that the DIF, DDF, and EOT factors could all load well on a higher-order ‘general alexithymia’ factor (loadings=0.84-0.91). That is, the DIF, DDF, and EOT subscale factors were all good indicators of a broader alexithymia construct. Some fit index values indicated that the higher-order version of this model was slightly worse fitting than the correlated version (see Table 2), indicating that the ‘general alexithymia’ factor did not perfectly account for the relationship between the DIF, DDF and EOT factors. However, because of the strong loadings of the DIF, DDF, and EOT factors on the higher-order ‘general alexithymia’ factor, and the adequate model fit according to RMSEA and SRMR, we judged there to be reasonable support for the high-order ‘general alexithymia’ factor (and hence the summing of all items into a total scale score).

In sum, the factor structure of the TAS-20 was best represented by three substantive correlated factors (DIF, DDF, EOT) and a reverse-scored item method factor. This solution produced mostly adequate levels of fit, but overall, fit index values appeared to be reduced by loading issues with six EOT items.

---Table 1---

---Table 2---

---Table 3---

**Factorial invariance**

We tested the invariance of the best fitting TAS-20 model (i.e., 3-factor correlated model+method) in the offender sample, comparing it with Preece et al.’s (2018a) community sample where this factor model had also been found to be the best solution. The measurement
(CFI=0.870) and structural (CFI=0.867) models were not substantially worse fitting than the configural model (CFI=0.875), indicating that the factor structure was invariant across the offender and community samples. That is, the factors, factor loadings, and covariances did not differ significantly between the samples.

**Internal Consistency Reliability**

Reliability coefficients were good for the total scale score, excellent for the DIF subscale, and acceptable for the DDF subscale (α > .70; see Table 1). The EOT subscale had low reliability (α=0.51), which was not improved by splitting it into importance of emotions (α=0.51) and pragmatic thinking (α=0.18) subscales, or by removing the reverse-scored items (α=0.45).

**Discussion**

Overall, most aspects of the TAS-20 performed well in this sample of male prisoners. Its factor structure was best represented by the traditional three alexithymia factors (DIF, DDF, EOT) with a reverse-scored method factor added. Nevertheless, six EOT items had low loadings, suggesting that they were not good indicators of EOT. Our findings are therefore similar to most older work in non-forensic samples (e.g., Bagby et al., 1994), and Parker et al.’s (2005) offender study, which also supported the presence of correlated DIF, DDF, and EOT factors within the TAS-20. More recent studies with non-offenders also highlight the presence of a prominent reverse-scored item method factor in the TAS-20 (e.g., Meganck et al., 2008; Preece et al., 2018a; Preece et al., 2020b; Tuliao et al., 2019), and thus our findings are highly consistent with these previous patterns. Indeed, we found this factor structure to be invariant across our offender and community samples. Previously, Parker et al. had speculated that certain TAS-20 items might be less relevant to incarcerated offenders, however, our factorial invariance results suggest that the TAS-20 items do function similarly across these sample types. Our results therefore suggest that this alexithymia construct holds
regardless of offender status, and for the first time, our results confirm statistically that TAS-20 scores can be meaningfully and confidently compared between offender and non-offender groups. Similarly, our finding that the DIF, DDF, and EOT subscale factors can all load on a higher-order ‘general alexithymia’ factor provides the first factor analytic support for the calculation of a TAS-20 total scale score in forensic populations as a marker of overall alexithymia.

In terms of reliability, we found that the TAS-20 total scale score and the DIF and DDF subscales had adequate internal consistency, but the EOT subscale did not. Similar to recent recommendations for non-forensic samples (e.g., Loas et al., 2017; Preece et al., 2018a), we therefore recommend against using the EOT subscale score in forensic settings, focusing instead only on the TAS-20 total scale score, and DIF and DDF subscale scores. For many forensic and research purposes, examining alexithymia at the overall (i.e., total scale score rather than subscale) level is likely to be sufficient (e.g., Bagby et al., 2007), and so our results provide good support for the utility of the TAS-20 in this respect. However, for more nuanced facet (i.e., subscale) level analysis, the TAS-20 appears to be somewhat more limited regardless of offender status, as EOT is conceptually an important component of alexithymia (Taylor et al., 1999) and forensic researchers are interested in it (e.g., Strickland et al., 2017). A possible solution would be a revision of the TAS-20 EOT items (e.g., Gignac et al., 2007; Meganck et al., 2008); another might be to use a multi-measure approach (Bagby et al., 2006). The Perth Alexithymia Questionnaire (PAQ; Preece et al., 2018b), was recently introduced and has an EOT subscale that has so far demonstrated high reliability in community and student populations (e.g., Greene et al., 2020; Preece et al., 2020b); it could now be evaluated in offender samples.

**Strengths and limitations of the study**

We think our study makes a strong contribution in terms of providing a psychometric
foundation for confident use of the TAS-20 in forensic settings, but it has limitations and therefore further research is necessary. Chiefly, to maximise sample size for our factor analyses, we treated our sample as a general offender sample, rather than splitting it into violent and non-violent offender groups. Offense type might impact performance on the TAS-20 and future research that examines larger samples of specific categories of offenders will therefore be beneficial. Similarly, the majority of our sample were Australian born and we did not include Aboriginal offenders. Future work could examine the extent to which our findings are generalisable across different countries, cultures, and TAS-20 language versions.

Conclusions

In sum, like in non-forensic samples, when used with incarcerated offenders the TAS-20 appears to have mostly adequate psychometric properties, but we recommend against using the EOT subscale score. The TAS-20 total scale score and DIF and DDF scores can be used confidently as markers of alexithymia in incarcerated offenders.
References


## Table 1

*Descriptive Statistics and Cronbach’s Alpha (α) for the TAS-20 in Incarcerated Offenders*

<table>
<thead>
<tr>
<th>Subscale/scale</th>
<th>$M$</th>
<th>$SD$</th>
<th>range</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIF</td>
<td>16.75</td>
<td>7.86</td>
<td>7-35</td>
<td>0.91</td>
</tr>
<tr>
<td>DDF</td>
<td>14.12</td>
<td>5.24</td>
<td>5-24</td>
<td>0.79</td>
</tr>
<tr>
<td>EOT</td>
<td>20.20</td>
<td>4.92</td>
<td>8-33</td>
<td>0.51</td>
</tr>
<tr>
<td>Total</td>
<td>51.08</td>
<td>15.17</td>
<td>20-89</td>
<td>0.88</td>
</tr>
</tbody>
</table>

*Note.* DIF = difficulty identifying feelings, DDF = difficulty describing feelings, EOT = externally orientated thinking.
Table 2

Goodness-of-Fit Index Values for the Different Confirmatory Factor Analysis Models of the TAS-20 in Incarcerated Offenders

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$ (df)</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA (90% CI)</th>
<th>SRMR</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-factor model</td>
<td>364.558 (170)</td>
<td>0.823</td>
<td>0.803</td>
<td>0.089 (0.076-0.101)</td>
<td>0.0751</td>
<td>444.558</td>
</tr>
<tr>
<td>2-factor correlated model</td>
<td>348.724 (169)</td>
<td>0.837</td>
<td>0.817</td>
<td>0.086 (0.073-0.098)</td>
<td>0.0764</td>
<td>430.724</td>
</tr>
<tr>
<td>3-factor correlated model</td>
<td>307.321 (167)</td>
<td>0.873</td>
<td>0.855</td>
<td>0.076 (0.063-0.089)</td>
<td>0.0728</td>
<td>393.321</td>
</tr>
<tr>
<td>4-factor correlated model</td>
<td>295.763 (164)</td>
<td>0.880</td>
<td>0.861</td>
<td>0.074 (0.061-0.088)</td>
<td>0.0669</td>
<td>387.763</td>
</tr>
<tr>
<td>3-factor correlated model + method</td>
<td>268.879 (162)</td>
<td>0.903</td>
<td>0.886</td>
<td>0.067 (0.053-0.081)</td>
<td>0.0638</td>
<td>364.879</td>
</tr>
<tr>
<td>3-factor higher-order model + method</td>
<td>275.924 (163)</td>
<td>0.898</td>
<td>0.881</td>
<td>0.069 (0.055-0.083)</td>
<td>0.0639</td>
<td>369.924</td>
</tr>
</tbody>
</table>

Note. For all models, $\chi^2 p < 0.001$. CFI = comparative fit index, TLI = Tucker-Lewis index, RMSEA = root mean square error of approximation, CI = confidence interval, SRMR = standardised root mean square residual, AIC = Akaike information criterion.
Table 3

**Standardised Factor loadings for the Different Confirmatory Factor Analysis Models of the TAS-20 in Incarcerated Offenders**

<table>
<thead>
<tr>
<th>Item/factor</th>
<th>1-factor model</th>
<th>2-factor correlated model</th>
<th>3-factor correlated model</th>
<th>4-factor correlated model</th>
<th>3-factor correlated model+method</th>
<th>3-factor higher-order model+method</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIF</td>
<td>[.902^\dagger]</td>
<td>[.674, .681]</td>
<td>[.686]</td>
<td>[.685]</td>
<td>[.687]</td>
<td>[.733]</td>
</tr>
<tr>
<td>1</td>
<td>[.648]</td>
<td>[.648]</td>
<td>[.663]</td>
<td>[.666]</td>
<td>[.663]</td>
<td>[.671]</td>
</tr>
<tr>
<td>3</td>
<td>[.812]</td>
<td>[.816]</td>
<td>[.828]</td>
<td>[.830]</td>
<td>[.828]</td>
<td>[.831]</td>
</tr>
<tr>
<td>6</td>
<td>[.702]</td>
<td>[.701]</td>
<td>[.720]</td>
<td>[.723]</td>
<td>[.720]</td>
<td>[.728]</td>
</tr>
<tr>
<td>7</td>
<td>[.834]</td>
<td>[.833]</td>
<td>[.836]</td>
<td>[.837]</td>
<td>[.836]</td>
<td>[.841]</td>
</tr>
<tr>
<td>9</td>
<td>[.803]</td>
<td>[.805]</td>
<td>[.813]</td>
<td>[.810]</td>
<td>[.813]</td>
<td>[.821]</td>
</tr>
<tr>
<td>13</td>
<td>[.784]</td>
<td>[.786]</td>
<td>[.789]</td>
<td>[.786]</td>
<td>[.789]</td>
<td>[.794]</td>
</tr>
<tr>
<td>DDF</td>
<td>[.912^\dagger]</td>
<td>[.711, .710]</td>
<td>[.798]</td>
<td>[.797]</td>
<td>[.799]</td>
<td>[.762]</td>
</tr>
<tr>
<td>2</td>
<td>[.442]</td>
<td>[.443]</td>
<td>[.471]</td>
<td>[.475]</td>
<td>[.459 (.202)]</td>
<td>[.457 (.215)]</td>
</tr>
<tr>
<td>4r</td>
<td>[.665]</td>
<td>[.665]</td>
<td>[.661]</td>
<td>[.660]</td>
<td>[.662]</td>
<td>[.664]</td>
</tr>
<tr>
<td>11</td>
<td>[.480]</td>
<td>[.478]</td>
<td>[.512]</td>
<td>[.511]</td>
<td>[.516]</td>
<td>[.517]</td>
</tr>
<tr>
<td>12</td>
<td>[.684]</td>
<td>[.677]</td>
<td>[.823]</td>
<td>[.823]</td>
<td>[.827]</td>
<td>[.813]</td>
</tr>
<tr>
<td>EOT or PR</td>
<td>[.836^\dagger]</td>
<td>[.022*, .001*]</td>
<td>[.011*]</td>
<td>[.025*]</td>
<td>[.016* (.067*)]</td>
<td>[.006 (.057*)]</td>
</tr>
<tr>
<td>5r</td>
<td>[.392]</td>
<td>[.442]</td>
<td>[.426]</td>
<td>[.449]</td>
<td>[.442]</td>
<td>[.470]</td>
</tr>
<tr>
<td>8</td>
<td>[.284]</td>
<td>[.269]</td>
<td>[.255]</td>
<td>[.314]</td>
<td>[.285]</td>
<td>[.310]</td>
</tr>
<tr>
<td>20</td>
<td>[.412]</td>
<td>[.527]</td>
<td>[.588]</td>
<td>[.608]</td>
<td>[.583]</td>
<td>[.545]</td>
</tr>
<tr>
<td>EOT or IM</td>
<td>[.348 (.276)]</td>
<td>[.338 (.283)]</td>
<td>[.348 (.276)]</td>
<td>[.338 (.283)]</td>
<td>[.348 (.276)]</td>
<td>[.338 (.283)]</td>
</tr>
<tr>
<td>10r</td>
<td>[.090*]</td>
<td>[.201*]</td>
<td>[.223]</td>
<td>[.189*]</td>
<td>[.238]</td>
<td>[.225]</td>
</tr>
<tr>
<td>15</td>
<td>[.133*]</td>
<td>[.292]</td>
<td>[.249]</td>
<td>[.302]</td>
<td>[.127* (.431)]</td>
<td>[.132 (.439)]</td>
</tr>
<tr>
<td>16</td>
<td>[.304]</td>
<td>[.523]</td>
<td>[.475]</td>
<td>[.526]</td>
<td>[.363 (.863*)]</td>
<td>[.375 (.840*)]</td>
</tr>
<tr>
<td>18r</td>
<td>[.304]</td>
<td>[.523]</td>
<td>[.475]</td>
<td>[.526]</td>
<td>[.363 (.863*)]</td>
<td>[.375 (.840*)]</td>
</tr>
<tr>
<td>19r</td>
<td>[.304]</td>
<td>[.523]</td>
<td>[.475]</td>
<td>[.526]</td>
<td>[.363 (.863*)]</td>
<td>[.375 (.840*)]</td>
</tr>
</tbody>
</table>

*Note. Factor loadings < .40 are boldface. For the method factor models, loadings inside brackets are loadings on the method factor, loadings outside the brackets loadings on the substantive alexithymia factor. DIF = difficulty identifying feelings, DDF = difficulty describing feelings, EOT = externally orientated thinking, IM = lack of importance of emotions, PR = pragmatic thinking, r = reverse-scored item.

^\dagger loading of first-order factor on higher-order ‘alexithymia’ factor.

*p > 0.05
Figure 1. The different confirmatory factor analysis models of the TAS-20. Squares represent observed variables (item numbers), ellipses represent latent factors. All items had an associated error term. Alexi = alexithymia, DIF = difficulty identifying feelings, DDF = difficulty describing feelings, EOT = externally orientated thinking, PR = pragmatic thinking, IM = lack of importance of emotions.